

CertiCrypt

Language-Based Cryptographic Proofs in Coq

Gilles Barthe^{1,2} Benjamin Grégoire^{1,3}
Santiago Zanella^{1,3}

¹Microsoft Research - INRIA Joint Centre, France



²IMDEA Software, Madrid, Spain



³INRIA Sophia Antipolis - Méditerranée, France



POPL 2009

What's wrong with cryptographic proofs?

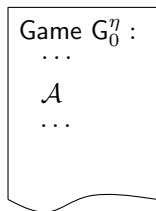
- *In our opinion, many proofs in cryptography have become essentially unverifiable. Our field may be approaching a crisis of rigor*
M. Bellare and P. Rogaway.
- *Do we have a problem with cryptographic proofs? Yes, we do [...] We generate more proofs than we carefully verify (and as a consequence some of our published proofs are incorrect)*
S. Halevi
- *Security proofs in cryptography may be organized as sequences of games [...] this can be a useful tool in taming the complexity of security proofs that might otherwise become so messy, complicated, and subtle as to be nearly impossible to verify*
V. Shoup

What's wrong with cryptographic proofs?

- *In our opinion, many proofs in cryptography have become essentially unverifiable. Our field may be approaching a crisis of rigor*
M. Bellare and P. Rogaway.
- *Do we have a problem with cryptographic proofs? Yes, we do [...] We generate more proofs than we carefully verify (and as a consequence some of our published proofs are incorrect)*
S. Halevi
- *Security proofs in cryptography may be organized as sequences of games [...] this can be a useful tool in taming the complexity of security proofs that might otherwise become so messy, complicated, and subtle as to be nearly impossible to verify*
V. Shoup

Game-based cryptographic proofs

Attack Game



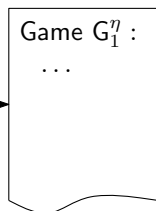
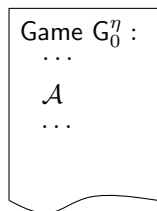
$\Pr_{G_0^\eta}[A_0]$

$$\Pr_{G_0^\eta}[A_0] \leq \epsilon(\eta)$$

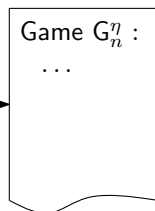
Security property

Game-based cryptographic proofs

Attack Game



...

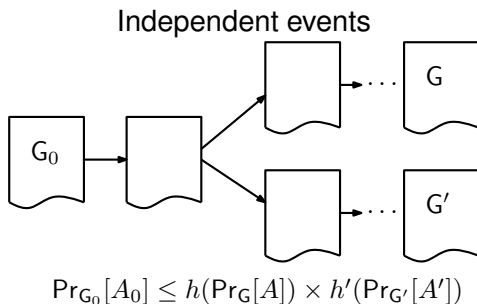


$$\Pr_{G_0^\eta}[A_0] \leq h_1(\Pr_{G_1^\eta}[A_1]) \leq \dots \leq h_n(\Pr_{G_n^\eta}[A_n])$$

$$\Pr_{G_0^\eta}[A_0] \leq h(\Pr_{G_n^\eta}[A_n]) \leq \epsilon(\eta)$$

Final Game

Game-based proofs: essence and problems

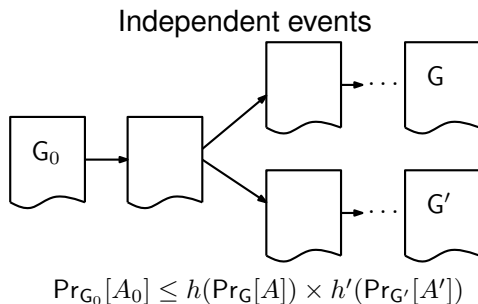


Essence: relate the probability of events in consecutive games

But,

- How do we represent games?
- What adversaries are *feasible*?
- How do we make a proof hold for any feasible adversary?

Game-based proofs: essence and problems



Essence: relate the probability of events in consecutive games

But,

- How do we represent games?
- What adversaries are *feasible*?
- How do we make a proof hold for any feasible adversary?

What if we represent games as programs?

Games	⇒	programs
Probability space	⇒	program denotation
Game transformations	⇒	program transformations
Generic adversary	⇒	unspecified procedure
Feasibility	⇒	Probabilistic Polynomial-Time

PWHILE: a probabilistic programming language

\mathcal{I}	$::=$	$\mathcal{V} \leftarrow \mathcal{E}$	assignment
		$\mathcal{V} \overset{\$}{\leftarrow} \mathcal{D}$	random sampling
		if \mathcal{E} then \mathcal{C} else \mathcal{C}	conditional
		while \mathcal{E} do \mathcal{C}	while loop
		$\mathcal{V} \leftarrow \mathcal{P}(\mathcal{E}, \dots, \mathcal{E})$	procedure call
\mathcal{C}	$::=$	nil	nop
		$\mathcal{I}; \mathcal{C}$	sequence

Measure monad: $M(X) \stackrel{\text{def}}{=} (X \rightarrow [0, 1]) \rightarrow [0, 1]$

$$\llbracket \cdot \rrbracket : \mathcal{C} \rightarrow \mathcal{M} \rightarrow M(\mathcal{M})$$

$$\llbracket x \overset{\$}{\leftarrow} \{0, 1\}; y \overset{\$}{\leftarrow} \{0, 1\} \rrbracket m =$$

Probability: $\text{Pr}_{G,m}[A] \stackrel{\text{def}}{=} \llbracket G \rrbracket m \mathbb{1}_A$

PWHILE: a probabilistic programming language

\mathcal{I}	::=	$\mathcal{V} \leftarrow \mathcal{E}$	assignment
		$\mathcal{V} \xleftarrow{\$} \mathcal{D}$	random sampling
		if \mathcal{E} then \mathcal{C} else \mathcal{C}	conditional
		while \mathcal{E} do \mathcal{C}	while loop
		$\mathcal{V} \leftarrow \mathcal{P}(\mathcal{E}, \dots, \mathcal{E})$	procedure call
\mathcal{C}	::=	nil	nop
		$\mathcal{I}; \mathcal{C}$	sequence

Measure monad: $M(X) \stackrel{\text{def}}{=} (X \rightarrow [0, 1]) \rightarrow [0, 1]$

$$\llbracket \cdot \rrbracket : \mathcal{C} \rightarrow \mathcal{M} \rightarrow M(\mathcal{M})$$

$$\begin{aligned} \llbracket x \xleftarrow{\$} \{0, 1\}; y \xleftarrow{\$} \{0, 1\} \rrbracket m f = & \\ & \frac{1}{4} f(m[0, 0/x, y]) \quad + \quad \frac{1}{4} f(m[0, 1/x, y]) \quad + \\ & \frac{1}{4} f(m[1, 0/x, y]) \quad + \quad \frac{1}{4} f(m[1, 1/x, y]) \end{aligned}$$

Probability: $\text{Pr}_{G,m}[A] \stackrel{\text{def}}{=} \llbracket G \rrbracket m \mathbb{1}_A$

PWHILE: a probabilistic programming language

\mathcal{I}	$::=$	$\mathcal{V} \leftarrow \mathcal{E}$	assignment
		$\mathcal{V} \xleftarrow{\$} \mathcal{D}$	random sampling
		if \mathcal{E} then \mathcal{C} else \mathcal{C}	conditional
		while \mathcal{E} do \mathcal{C}	while loop
		$\mathcal{V} \leftarrow \mathcal{P}(\mathcal{E}, \dots, \mathcal{E})$	procedure call
\mathcal{C}	$::=$	nil	nop
		$\mathcal{I}; \mathcal{C}$	sequence

Measure monad: $M(X) \stackrel{\text{def}}{=} (X \rightarrow [0, 1]) \rightarrow [0, 1]$

$$\llbracket \cdot \rrbracket : \mathcal{C} \rightarrow \mathcal{M} \rightarrow M(\mathcal{M})$$

$$\begin{aligned} \llbracket x \xleftarrow{\$} \{0, 1\}; y \xleftarrow{\$} \{0, 1\} \rrbracket m \mathbb{1}_{x \neq y} = & \\ & \frac{1}{4} \mathbb{1}_{x \neq y}(m[0, 0/x, y]) + \frac{1}{4} \mathbb{1}_{x \neq y}(m[0, 1/x, y]) + \\ & \frac{1}{4} \mathbb{1}_{x \neq y}(m[1, 0/x, y]) + \frac{1}{4} \mathbb{1}_{x \neq y}(m[1, 1/x, y]) \end{aligned}$$

Probability: $\text{Pr}_{G,m}[A] \stackrel{\text{def}}{=} \llbracket G \rrbracket m \mathbb{1}_A$

PWHILE: a probabilistic programming language

\mathcal{I}	::=	$\mathcal{V} \leftarrow \mathcal{E}$	assignment
		$\mathcal{V} \xleftarrow{\$} \mathcal{D}$	random sampling
		if \mathcal{E} then \mathcal{C} else \mathcal{C}	conditional
		while \mathcal{E} do \mathcal{C}	while loop
		$\mathcal{V} \leftarrow \mathcal{P}(\mathcal{E}, \dots, \mathcal{E})$	procedure call
\mathcal{C}	::=	nil	nop
		$\mathcal{I}; \mathcal{C}$	sequence

Measure monad: $M(X) \stackrel{\text{def}}{=} (X \rightarrow [0, 1]) \rightarrow [0, 1]$

$$\llbracket \cdot \rrbracket : \mathcal{C} \rightarrow \mathcal{M} \rightarrow M(\mathcal{M})$$

$$\llbracket x \xleftarrow{\$} \{0, 1\}; y \xleftarrow{\$} \{0, 1\} \rrbracket m \mathbb{1}_{x \neq y} =$$

0	+	$\frac{1}{4}$	+
$\frac{1}{4}$	+	0	

Probability: $\Pr_{G,m}[A] \stackrel{\text{def}}{=} \llbracket G \rrbracket m \mathbb{1}_A$

PWHILE: a probabilistic programming language

\mathcal{I}	::=	$\mathcal{V} \leftarrow \mathcal{E}$	assignment
		$\mathcal{V} \overset{\$}{\leftarrow} \mathcal{D}$	random sampling
		if \mathcal{E} then \mathcal{C} else \mathcal{C}	conditional
		while \mathcal{E} do \mathcal{C}	while loop
		$\mathcal{V} \leftarrow \mathcal{P}(\mathcal{E}, \dots, \mathcal{E})$	procedure call
\mathcal{C}	::=	nil	nop
		$\mathcal{I}; \mathcal{C}$	sequence

Measure monad: $M(X) \stackrel{\text{def}}{=} (X \rightarrow [0, 1]) \rightarrow [0, 1]$

$$\llbracket \cdot \rrbracket : \mathcal{C} \rightarrow \mathcal{M} \rightarrow M(\mathcal{M})$$

$$\llbracket x \overset{\$}{\leftarrow} \{0, 1\}; y \overset{\$}{\leftarrow} \{0, 1\} \rrbracket m \mathbb{1}_{x \neq y} = \frac{1}{2}$$

Probability: $\text{Pr}_{G,m}[A] \stackrel{\text{def}}{=} \llbracket G \rrbracket m \mathbb{1}_A$

Untyped vs. typed language

- 1st attempt: untyped language, lots of problems
 - No guarantee that programs are well-typed
 - Had to deal with ill-typed programs
- 2nd attempt: typed language (dependently typed syntax!)
 - Programs are well-typed by construction

Inductive $\mathcal{I} : \text{Type} :=$

| Assign : $\forall t, \mathcal{V}_t \rightarrow \mathcal{E}_t \rightarrow \mathcal{I}$

| Rand : $\forall t, \mathcal{V}_t \rightarrow \mathcal{D}_t \rightarrow \mathcal{I}$

| Cond : $\mathcal{E}_{\text{Bool}} \rightarrow \mathcal{C} \rightarrow \mathcal{C} \rightarrow \mathcal{I}$

| While : $\mathcal{E}_{\text{Bool}} \rightarrow \mathcal{C} \rightarrow \mathcal{I}$

| Call : $\forall l t, \mathcal{P}_{(l,t)} \rightarrow \mathcal{V}_t \rightarrow \mathcal{E}_l^* \rightarrow \mathcal{I}$

where $\mathcal{C} := \mathcal{I}^*$.

Parametrized semantics: $[[\cdot]] : \forall \eta, \mathcal{C} \rightarrow \mathcal{M} \rightarrow M(\mathcal{M})$

Untyped vs. typed language

- 1st attempt: untyped language, lots of problems
 - No guarantee that programs are well-typed
 - Had to deal with ill-typed programs
- 2nd attempt: typed language (dependently typed syntax!)
 - Programs are well-typed by construction

Inductive $\mathcal{I} : \text{Type} :=$

| Assign : $\forall t, \mathcal{V}_t \rightarrow \mathcal{E}_t \rightarrow \mathcal{I}$

| Rand : $\forall t, \mathcal{V}_t \rightarrow \mathcal{D}_t \rightarrow \mathcal{I}$

| Cond : $\mathcal{E}_{\text{Bool}} \rightarrow \mathcal{C} \rightarrow \mathcal{C} \rightarrow \mathcal{I}$

| While : $\mathcal{E}_{\text{Bool}} \rightarrow \mathcal{C} \rightarrow \mathcal{I}$

| Call : $\forall l t, \mathcal{P}_{(l,t)} \rightarrow \mathcal{V}_t \rightarrow \mathcal{E}_l^* \rightarrow \mathcal{I}$

where $\mathcal{C} := \mathcal{I}^*$.

Parametrized semantics: $[[\cdot]] : \forall \eta, \mathcal{C} \rightarrow \mathcal{M} \rightarrow M(\mathcal{M})$

Characterizing feasible adversaries

A cost model for reasoning about program complexity

$$\llbracket \cdot \rrbracket' : \forall \eta, \mathcal{C} \rightarrow (\mathcal{M} \times \mathbb{N}) \rightarrow M(\mathcal{M} \times \mathbb{N})$$

Non-intrusive:

$$\llbracket G \rrbracket m = \text{bind} (\llbracket G \rrbracket' (m, 0)) (\lambda mn. \text{unit} (\text{fst } mn))$$

A program G runs in probabilistic polynomial time if:

- It terminates with probability 1 (i.e. $\forall m, \Pr_{G,m}[\text{true}] = 1$)
- There exists a polynomial $p(\cdot)$ s.t. if (m', n) is reachable with positive probability, then $n \leq p(\eta)$

Characterizing feasible adversaries

A cost model for reasoning about program complexity

$$\llbracket \cdot \rrbracket' : \forall \eta, \mathcal{C} \rightarrow (\mathcal{M} \times \mathbb{N}) \rightarrow M(\mathcal{M} \times \mathbb{N})$$

Non-intrusive:

$$\llbracket G \rrbracket m = \text{bind} (\llbracket G \rrbracket' (m, 0)) (\lambda mn. \text{unit} (\text{fst } mn))$$

A program G runs in probabilistic polynomial time if:

- It terminates with probability 1 (i.e. $\forall m, \Pr_{G,m}[\text{true}] = 1$)
- There exists a polynomial $p(\cdot)$ s.t. if (m', n) is reachable with positive probability, then $n \leq p(\eta)$

Program equivalence

Definition (Observational equivalence)

$$f =_X g \stackrel{\text{def}}{=} \forall m_1 m_2, m_1(X) = m_2(X) \implies f m_1 = g m_2$$

$$\models G_1 \simeq_O^I G_2 \stackrel{\text{def}}{=} \forall m_1 m_2 f g, m_1(I) = m_2(I) \wedge f =_O g \implies \llbracket G_1 \rrbracket m_1 f = \llbracket G_2 \rrbracket m_2 g$$

Generalizes information flow security.

But is not general enough...

???

$$\models \text{if } x = 0 \text{ then } y \leftarrow x \text{ else } y \leftarrow 1 \simeq_{\{x\}}^{\{x,y\}} \text{if } x = 0 \text{ then } y \leftarrow 0 \text{ else } y \leftarrow 1$$

Program equivalence

Definition (Observational equivalence)

$$f =_X g \stackrel{\text{def}}{=} \forall m_1 m_2, m_1(X) = m_2(X) \implies f m_1 = g m_2$$
$$\models G_1 \simeq_O^I G_2 \stackrel{\text{def}}{=} \forall m_1 m_2 f g, m_1(I) = m_2(I) \wedge f =_O g \implies \llbracket G_1 \rrbracket m_1 f = \llbracket G_2 \rrbracket m_2 g$$

Generalizes information flow security.
But is not general enough...

???

$$\models \text{if } x = 0 \text{ then } y \leftarrow x \text{ else } y \leftarrow 1 \simeq_{\{x,y\}}^{\{x\}} \text{if } x = 0 \text{ then } y \leftarrow 0 \text{ else } y \leftarrow 1$$

Program equivalence

Definition (Observational equivalence, generalization)

$$\models G_1 \sim G_2 : \Psi \Rightarrow \Phi \stackrel{\text{def}}{=}$$

$$\forall m_1 m_2. m_1 \Psi m_2 \Rightarrow \llbracket G_1 \rrbracket m_1 \sim_\Phi \llbracket G_2 \rrbracket m_2$$

Where \sim_Φ is the lifting of relation Φ from memories to distributions.

$$(x = 0) \sim_{\{x\}} (x = 0)$$

$$\models y \leftarrow x \sim y \leftarrow 0 := \{x\} \wedge (x = 0) \langle 1 \rangle \Rightarrow = \{x, y\}$$

$$\models y \leftarrow 1 \sim y \leftarrow 1 := \{x\} \wedge (x \neq 0) \langle 1 \rangle \Rightarrow = \{x, y\}$$

$$\text{if } x = 0 \text{ then } y \leftarrow x \text{ else } y \leftarrow 1 \sim$$

$$\text{if } x = 0 \text{ then } y \leftarrow 0 \text{ else } y \leftarrow 1 := \{x\} \Rightarrow = \{x, y\}$$

From program equivalence to probability

Let A be an event that depends only on variables in O

To prove $\Pr_{G_1, m_1}[A] = \Pr_{G_2, m_2}[A]$ it suffices to show

- $\models G_1 \simeq_O^I G_2$
- $m_1 =_I m_2$

Proving program equivalence

Goal

$$\models G_1 \simeq'_O G_2$$

A Relational Hoare Logic

$$\frac{\models c_1 \sim c_2 : \Phi \Rightarrow \Phi' \quad \models c'_1 \sim c'_2 : \Phi' \Rightarrow \Phi''}{\models c_1; c'_1 \sim c_2; c'_2 : \Phi \Rightarrow \Phi''} \text{[R-Seq]}$$

...

Proving program equivalence

Goal

$$\vDash G_1 \simeq_O^I G_2$$

Mechanized program transformations

- Transformation: $T(G_1, G_2, I, O) = (G'_1, G'_2, I', O')$
- Soundness theorem

$$\frac{T(G_1, G_2, I, O) = (G'_1, G'_2, I', O') \quad \vDash G'_1 \simeq_{O'}^{I'} G'_2}{\vDash G_1 \simeq_O^I G_2}$$

- Reflection-based Coq tactic

Proving program equivalence

Goal

$$\models G_1 \simeq_O^I G_2$$

Mechanized program transformations

- Dead code elimination
- Constant folding and propagation
- Procedure call inlining
- Instruction reordering
- Common suffix/prefix elimination

Proving program equivalence

Goal

$$\models G_1 \simeq_O^I G_2$$

A semi-decision procedure for self-equivalence

- Does $\models G \simeq_O^I G$ hold?
- Analyze dependencies to compute I' s.t. $\models G \simeq_O^{I'} G$
- Check that $I' \subseteq I$

Example

Game ElGamal₀ :

```
x ←$ ℤq; y ←$ ℤq;
(m0, m1) ← A(gx);
b ←$ {0, 1};
ζ ← gxy × mb;
b' ← A'(gx, gy, ζ);
d ← b = b'
```

```
inline_r B;
ep;
deadcode;
eqobs_in
```

$\simeq_{\emptyset}^{\{d\}}$

Game DDH₀ :

```
x ←$ ℤq;
y ←$ ℤq;
d ← B(gx, gy, gxy)
```

Procedure B(α, β, γ) :

```
(m0, m1) ← A(α);
b ←$ {0, 1};
b' ← A'(α, β, γ × mb);
return b = b'
```

The Fundamental Lemma of Game-Playing

Fundamental lemma

If two games G_1 and G_2 behave identically in an initial memory m unless a failure event A fires, then

$$|\Pr_{G_1,m}[A] - \Pr_{G_2,m}[A]| \leq \Pr_{G_{1,2}}[F]$$

The Fundamental Lemma of Game-Playing

Game G_1 :

...

bad \leftarrow true; c_1

...

Game G_2 :

...

bad \leftarrow true; c_2

...

- $\Pr_{G_1,m}[A \wedge \neg \text{bad}] = \Pr_{G_2,m}[A \wedge \neg \text{bad}]$
- $\Pr_{G_1,m}[\text{bad}] = \Pr_{G_2,m}[\text{bad}]$

Corollary

$$|\Pr_{G_1,m}[A] - \Pr_{G_2,m}[A]| \leq \Pr_{G_1,2}[\text{bad}]$$

Wrapping up

Contributions

- Formal semantics of a probabilistic programming language
- Characterization of probabilistic polynomial-time programs
- A Probabilistic Relational Hoare logic
- Mechanization of common program transformations
- Formalized emblematic proofs: ElGamal, FDH, OAEP

Perspectives

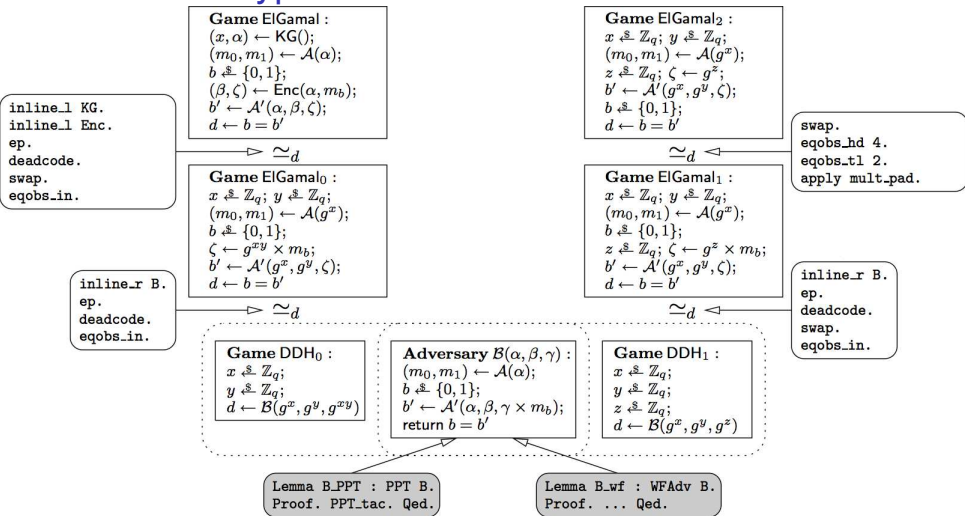
- Overwhelming number of applications: IB, ZK proofs, ...
- Computational soundness of symbolic methods and information flow type systems
- Verification of randomized algorithms

Some statistics

- 6 persons involved
- CertiCrypt: 30,000 lines of Coq, 48 man-months
- Full Domain Hash: 2,500 lines of Coq, 4 man-months
(for a person without experience in CertiCrypt)

Questions

ElGamal encryption



$$\left| \Pr_{\text{ElGamal}}[b = b'] - \frac{1}{2} \right| = |\Pr_{\text{DDH}_0}[d] - \Pr_{\text{DDH}_1}[d]|$$

Observational equivalence

$$\models G_1 \sim G_2 : \Psi \Rightarrow \Phi \stackrel{\text{def}}{=} m_1 \Psi m_2 \Rightarrow \llbracket G_1 \rrbracket m_1 \sim_{\Phi} \llbracket G_2 \rrbracket m_2$$

Lifting

$$\text{range } P \mu \stackrel{\text{def}}{=} \forall f, (\forall a, P a \Rightarrow f a = 0) \Rightarrow \mu f = 0$$

$$\mu_1 \sim_{\Phi} \mu_2 \stackrel{\text{def}}{=} \exists \mu, \pi_1(\mu) = \mu_1 \wedge \pi_2(\mu) = \mu_2 \wedge \text{range } \Phi \mu$$

Small-step semantics

$$(\text{nil}, m, []) \rightsquigarrow \text{unit}(\text{nil}, m, [])$$
$$(\text{nil}, m, (x, e, c, l) :: F) \rightsquigarrow \text{unit}(c, (l, m.\text{glob})\{\llbracket e \rrbracket m/x\}, F)$$
$$(x \leftarrow p(\vec{e}); c, m, F) \rightsquigarrow \text{unit}(E(p).\text{body}, (\emptyset\{\llbracket \vec{e} \rrbracket m/E(p).\text{params}\},$$
$$(\text{if } e \text{ then } c_1 \text{ else } c_2; c, m, F) \rightsquigarrow \text{unit}(c_1; c, m, F)$$
$$\text{if } \llbracket e \rrbracket m = \text{true}$$
$$(\text{if } e \text{ then } c_1 \text{ else } c_2; c, m, F) \rightsquigarrow \text{unit}(c_2; c, m, F)$$
$$\text{if } \llbracket e \rrbracket m = \text{false}$$
$$(\text{while } e \text{ do } c; c', m, F) \rightsquigarrow \text{unit}(c; \text{while } e \text{ do } c; c', m, F)$$
$$\text{if } \llbracket e \rrbracket m = \text{true}$$
$$(\text{while } e \text{ do } c; c', m, F) \rightsquigarrow \text{unit}(c', m, F)$$
$$\text{if } \llbracket e \rrbracket m = \text{false}$$
$$(x \leftarrow e; c, m, F) \rightsquigarrow \text{unit}(c, m\{\llbracket e \rrbracket m/x\}, F)$$
$$(x \leftarrow \underline{\$} d; c, m, F) \rightsquigarrow \text{bind}(\llbracket d \rrbracket m)(\lambda v. \text{unit}(c, m\{v/x\}, F))$$

Denotation

$$\begin{aligned} \llbracket S \rrbracket_0 &\stackrel{\text{def}}{=} \text{unit } S & \llbracket S \rrbracket_{n+1} &\stackrel{\text{def}}{=} \text{bind } \llbracket S \rrbracket_n \llbracket \cdot \rrbracket^1 \\ \llbracket c \rrbracket m : M(\mathcal{M}) &\stackrel{\text{def}}{=} \lambda f. \sup \{ \llbracket (c, m, [\]) \rrbracket_n f \mid \text{final} \mid n \in \mathbb{N} \} \end{aligned}$$